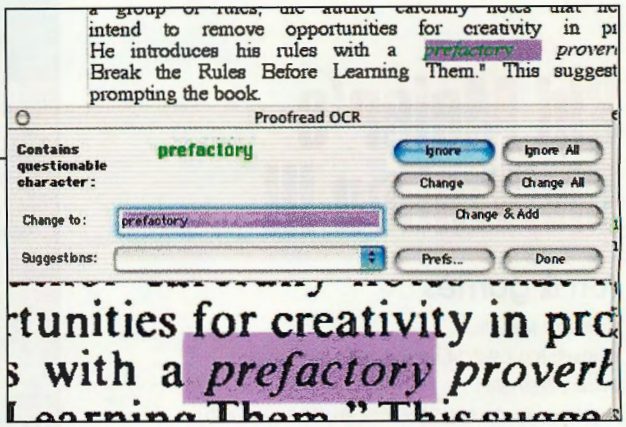
OmniPage Pro X

productivity

Company: Scansoft
Contact: 978-977-2000, www.scansoft.com
Price: $499 (SRP), $149.99 (upgrade from any OCR program)
Requirements: G3 or later, Mac OS 9.x or 10.1 or later, 64MB RAM (128MB for OS X), scanner



OmniPage can do a whole lot more than just recognize characters: The Proofreader feature also detects misspelled words.

It's been three years since Scansoft last updated OmniPage, but OmniPage Pro X, the latest generation in OCR (optical character recognition) software, was worth the wait. OCR is the process of turning a graphical scan into editable text.

OmniPage's OCR accuracy is better than ever—it's not perfect, but it's usually very good, even on pages with complex formatting. The program highlights words it's unsure of, and the proofreading function makes it easy to locate and correct errors. In addition to offering a built-in dictionary, OmniPage supports user-created dictionaries—you can even import custom dictionaries from Microsoft Word. OmniPage also has improved table manners compared with the previous version's; the program handily recognized our table-bound data.

Once it has scanned, recognized, and proofread your pages, OmniPage Pro X can export the text in a variety of formats—with varying degrees of success. Microsoft Word files look good, retaining the layout of the original page. But HTML and RTF output often leaves much to be desired: Misplaced graphics and too-small text plagued our test pages.

The application can also export documents as Portable Document Format (PDF) files. You can choose from several flavors of PDF, including Images Only; Image On Text, which retains the exact look of the original document but has searchable text; and Normal, which reproduces the look of the original page using text instead of graphics, creating a much smaller file. Unfortunately the text is often irregular, with distracting changes in fonts and baseline height.

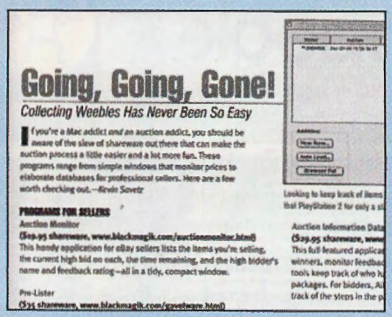
OmniPage is compatible with automatic document feeders for scanning multiple-page documents. But OmniPage doesn't include scanner drivers, so you'll need to install it over a working scanner setup, with a TWAIN driver or a Photoshop-scanner plug-in. OmniPage can also suck text directly out of PDF files (even read-only ones), so you technically don't even *need* a scanner.

Given the sad state of the OS X–compatible scanner market, OmniPage's OS X support is more a badge of honor than a functional benefit, but with the right scanner—or with good old Mac OS 9—OmniPage is a capable OCR program, especially if you don't need fancily formatted output.—*Kevin Savetz*

good news: Accurate OCR with good correction tools. PDF support. Good support for tables. **bad news:** Output quality is poor for some file types, especially on complex pages.

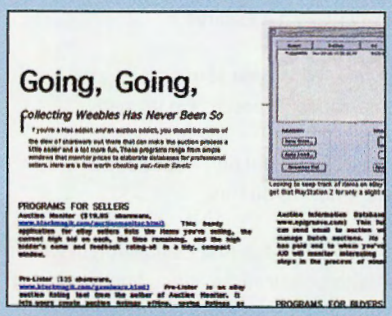
The Proof's in the Outputting

OCR is only half the battle: The other half lies in what happens when you export the pages. OmniPage Pro X renders pages very differently depending on what file format you use.

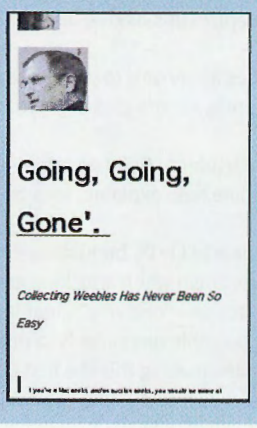


Here's the scan OmniPage started with.

When saved as a PDF file with images, the document looks remarkably like the original, but the text is slightly skewed.



The document looks good but isn't quite perfect when exported as a Microsoft Word file.



When saved as a Web page, documents with complex layouts turn to mush, with misplaced graphics and tiny, unreadable body text.